STUDENT ACHIEVEMENT PREDICTION MODEL BASED ON TLBO-BP

Huakun Hou, Xiaoke Deng, Zekun Lu and Linbo Zhai*

Department of Information Engineering, Shandong Normal University, Jinan, China

ABSTRACT

Today, educational prediction analysis has become an important tool for educational institutions to analyze students, and the performance of students in educational prediction analysis is a critical feedback. However, there are serious challenges in dealing with multi-factor datasets to improve the convergence and accuracy of predicting student performance. Therefore, this article comprehensively analyzes machine learning technologies, analyzes the community activities of middle-term students, and explores the impact of community activities on middle-aged students' practical skills. This paper uses Random Forest and Pearson Correlation Coefficient to analyze the relevance of the impact of community activities on student hands-on performance and use Teaching and Learning Algorithms (TLBO) to optimize Back propagation (BP) neural networks applied to predict students' future trends. The results show that the TLBO-BP model can more accurately predict dynamic changes in student performance and predict high accuracy in simple patterns.

KEYWORDS

Grade prediction, Feature analysis, Teaching and learning optimization algorithms, BP neural network

1. INTRODUCTION

As vocational education flourishes, its exposed problems become increasingly evident. The most promising concern of society is how to cultivate middle-term students that meet the requirements of social development so that middle-term students can get a skill length. At present, in middle-term schools, the process of competing for professional skills in a community-apprenticeship model has been formed.

A new educational teaching model has emerged in some middle-term schools. This model encourages all students to participate in community activities according to their interests. The school arranges a dedicated time for community activities and filters teachers to lead students, making community activities a form of educational organization that extends and complements classroom teaching. By evaluating the selection of students with excellent achievements within the community, the way in which skill competitions enable senior skills to be learned for all students escapes the trouble of elite education.

Education Data Mining (EDM) and Machine Learning (ML) methods have become common tools to help students improve learning outcomes. Using EDM and ML techniques and methods to analyze large amounts of learning data has produced interesting and explainable information about students.

David C. Wyld et al. (Eds): CMLA, GRAPH-HOC, CIoT, DMSE, ArIT, WiMoNe, CSEIT – 2023
pp. 13-23, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.131002

machine learning has been used to address multiple educational fields, including student performance, exit forecasting, academic early warning systems and course selection [4]. In previous studies, there were some studies on predicting student performance. For example, in [5], authors predicted student performance through BP neural networks, and in [6], author optimized BP neural networks by combining group intelligence algorithms but ignored the analysis of relevant factors.

Although there are several studies on predicting student grades, such as [1], [2],However, there was a lack of analysis of the reasons for the achievements. In this article, we focus on the prediction of student outcomes by observing student community activities. For this purpose, we initially collect student characteristics data from community game courses. In community activities, learners can continuously improve their practical abilities through practical hands-on operations. Then, by selecting high-level characteristics by Pearson Correlation Coefficient and Random Forest Classification algorithms, removing irrelevant and unrelated student features, the performance of students is predicted by BP Neural Network Optimization (TLBO-BP).

Our main contributions are described as follows:

1. This paper explores the implications of community activity models for the development of middle-term students.
2. We use random forest classification algorithms with Pearson Correlation Coefficient to determine the best selection characteristics and reduce the complexity of algorithms for student performance predictions.
3. This paper combines TLBO algorithms with neural network learning to predict future performance trends for students.
4. Simulations prove that our algorithm achieves greater advantages over BP neural networks, ACO-BP predictions, and GA-BP predictions.

The rest of this article is organized as follows. Section 2 presents the relevant work, and Section 3 outlines our experimental algorithms in detail. Data descriptions, evaluation methods and results are discussed in Section 4. Finally, it is summarized in Section 5.

2. RELATED WORK

In recent years, a variety of machine-based learning approaches have been proposed to improve student learning. These approaches solve different learning problems, such as predicting how well students perform by analyzing their interactions with the system, or early warning approaches to inform the students and course instructors, or recommending learning materials that students can timely complete. Among all these studies, several research aimed at predicting the students' final grades for a course. Students' grades were predicted using different algorithms, methods and techniques, such as using matrix factorization (MF) algorithms [7], a neural network, a decision tree (DT) [8], an RF [7]. These algorithms and methods can be divided into two groups. The methods that use only students' performance data for the prediction task and the ones that use also courses' or students' metadata, such as students' demo-graphic data or behavioral patterns [10] .

Character selection is the selection of a number of representative subsets of features from the original character concentration, and the classification or regression model built on that character subset achieves approximate or even better predictive accuracy than before the character selection. Character selection not only improves the space and time efficiency of the application of the algorithm to avoid "dimensional disasters", but also avoids the over-adaptation of the problem to a certain extent.

Various performance prediction models are widely developed and applied to help students at risk of failure. However, due to the unbalanced nature of data sets that lead to deviant results, it is challenging to establish high-precision predictive models. In [11], the authors predicted student performance by supporting vector regression (SVR), but it relies heavily on correct data representation. In [12], the authors predicted student outcomes through Pure Bayes (NB). Compared to vector regression prediction, the BP neural network has strong nonlinear mapping capabilities. As the scope of applications gradually expanded, BP neural networks also exposed more and more shortcomings and deficiencies, such as local minimization problems and the slow convergence of BP neural network algorithms. To address these problems, the use of a genetic algorithm (GA) and an ant group (ACO) is proposed in [13] and [14] to optimize BP neural networks in order to solve the problem of local minimum values in the presence of BP neural networks. To increase the convergence rate of BP neural networks, we use Teaching-Based Optimization Algorithms (TLBO) to optimize BP Neural Networks.

3. Framework Work

3.1. Analysis of Relevance and Importance

Character selection is often an important step in the application of machine learning methods. Descriptions of modern datasets often contain too many variables to make it difficult to build actual models. Usually, most of these variables are not related to the final outcome.

Random Forest (RF) is a typical feature selection algorithm that incorporates a characteristic importance evaluation mechanism while maintaining a better selection effect. The random forest classification algorithm by means of characteristic random substitution before and after error analysis calculates each characteristic's importance score [15].

This paper uses random forests to analyze educational data and determine relevance by comparing real characteristics with the relevance of random probes. The method of measuring the importance of random forest characteristics is the random substitution of each characteristic. Each characteristic is rated according to the variable value of the error rate, and thus a characteristic importance score is obtained.

As mentioned above, the main purpose of this study is to analyze the relationship between community activity outcomes and student academic achievements and to predict students' academic results through community activities in order to highlight the areas that need to be focused on when designing a more informative curriculum.

The characteristic importance is sorted by random forest, and the relevance is then further sorted according to the size of the correlation between the characteristics. Its relevance is measured using Pearson Correlation Coefficient ρ_{XY}. In formula (1), x_i is the i-th indicator variable, ρ_{xi},ρ_{xj} is the Pearson Correlation Coefficient between the *i*-th indicator variable and the *j*-th metric variable. Then the Pearson Correlation Coefficient between the populations of the 2 indicator variables is defined as:

$$\rho_{x_i,x_j} = \frac{cov(X_i,X_j)}{\delta_{X_i}\delta_{X_j}} = \frac{E\left[\left(X_i - u_{X_i}\right)\left(X_j - u_{X_j}\right)\right]}{\delta_{X_i}\delta_{X_j}} \tag{1}$$

where $cov(X_i, X_j)$ is the covariance between the *i*-th index variable and the *j*-th index variable, δ_{Xi} is the population standard deviation of the *i*-th index variable, and u_{Xi} is the population mean of the *i*-th index variable.

Table 1 Correlation coefficient meaning

Variable	Description
0.8-1.0	Extremely strongly correlated
0.6-0.8	Strong correlation
0.4-0.6	Moderately relevant
0.2-0.4	Weakly correlated
0.0-0.2	Very weak or no correlation

3.2. Deep BP Neural Network Algorithm Optimized by TLBO

The BP network consists of an input layer, a hidden layer, and an output layer, shown in Fig 1. BP network has high nonlinearity and strong generalization ability [16]. In order to improve the convergence speed of BP neural networks, we propose a method to optimize BP neural networks using Teaching-Learning-Based Optimization Algorithm.

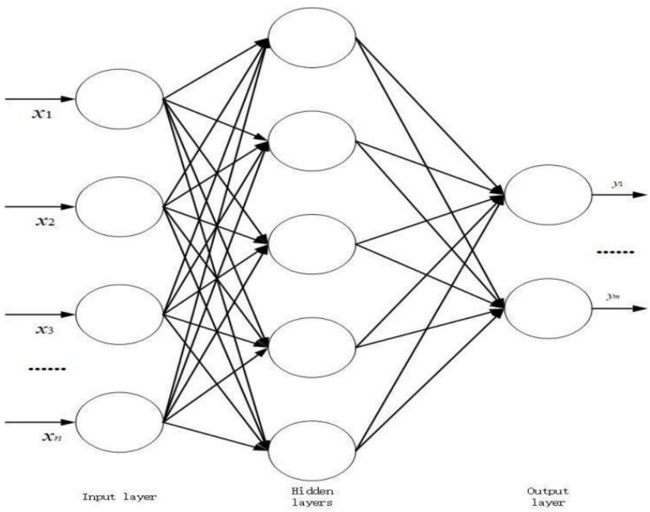


Fig. 1 BP neural network structure diagram

Teaching-Learning-Based Optimization Algorithm (TLBO) is a crowd-based, inspirational, random group intelligence algorithm [17]. Similar to other evolutionary algorithms, the approach also has iterative processes. The process is divided into two stages, each of which performs its own optimization. Compared to other algorithms, the main advantages of teaching optimization are the simplicity of the concept, the reduced number of overruns, and the rapid convergence.

In teaching optimization algorithms, students are viewed as search points distributed in the decision variable space, where the best students are defined as the teacher of the class. Unlike traditional evolutionary algorithms and group intelligence algorithms, the iterative evolutionary process of teaching optimization algorithms includes the teaching phase and the learning phase. In order to improve the average level of the class, students will improve their own level by learning from teachers. Then, in the study stage, they will increase their level by interacting with another student of random choice.

The principle of the TLBO algorithm derives from the relationship between teachers and students in the educational classroom environment [18], the effect of teachers on learners or students, and the interaction between learners and their interactions. The teacher phase and the learner phase are the two main parts of the algorithm, known respectively as the teacher phase and the learner phase.

This paper uses equal root error (RMSE) to evaluate the performance of the model, and the RMSE formula is defined as follows:

$$R = \sqrt{\frac{1}{T}\sum_{s=1}^{T}(x_s - \hat{x}_s)^2} \tag{2}$$

where x_s and $\widehat{x_s}$ respectively represent the *s*-th original value and predicted value in the sequence, and T represents the length of the time series.

All connection values and thresholds in the BP network are encoded as individuals, and then the N individuals form a cluster. BP neural networks often lead to training speed instability and local low values. Therefore, the results of training are not optimal. Because the TLBO algorithm has strong global search capabilities and convergence speed characteristics, TLBO is initialized to the deep BP neural network, i.e., to optimize the neuronal networks by TLBO[19]. The TLBO algorithm finds the optimal individual to allocate the initial values of the neurons and threads. The TLBO-BP algorithm is as follows:

- Student initialization

Remember that the decision variable dimension is *D*, and the *i*-th student (search point) is represented as $X_i = (x_{i1}, x_{i2}, ..., x_{iD}$, f_{xi} is represented as a fitness function, and the goal is to be as close as possible to the minimum of the function. *N* represents the total number of students. The *i*-th student $X_i = (x_{i1}, x_{i2}, ..., x_{iD}$ in the class can be initialized as:

$$X_{ij} = x_j^{min} + rand \times (x_j^{min} - x_j^{max}) \tag{3}$$

where $x^{max}{}_j$ and $x^{min}{}_j$ are the upper and lower bounds of the *j*-dimensional decision variable, and rand is a random number between [0,1].

- Teaching phase

During the teaching phase, simulated students improve themselves by learning about the difference between the teacher and the class average. For the *i*-th learnerx_j in the class, the update mechanism is expressed as follows:

$$newX_i = X_i + rand \times (Teacher - Mean) \tag{4}$$

where $newX_i$ is the new value of student X_i, and teacher is the student of the current optimal value, $\mathrm{Mean} = 1/\mathrm{N} \times \sum_1^{\mathrm{N}} \mathrm{X_i}$ is the average of the class. After the teaching phase, the student's value will take the lesser between their original fitness and the new fitness and then move on to the learning phase.

- Learning phase

In the learning phase, students learn in a simulated way of discussing, presenting and communicating with each other to improve themselves. For learner X_i, the update formula is:

$$newX_i = f(x) = \begin{cases} X_i + rand \times (X_i - X_k), \ if \ f(X_i) < f(X_k) \\ X_i + rand \times (X_k - X_i), \ otherwise \end{cases} \tag{5}$$

where $newX_i$ is the new value of student X_i, X_k is another student randomly selected from the class who is different from X_i, and $f(X_i)$ and $f(X_k)$ are the fitness of students X_i and X_k respectively.

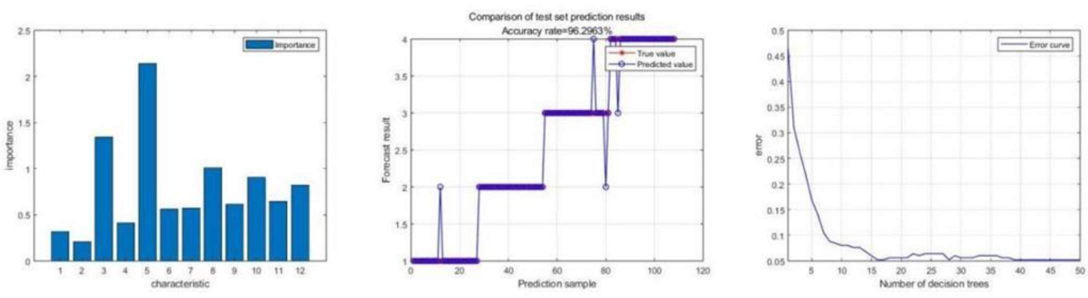
4. Experimental Result

In this section, we first introduce the experimental setup. The proposed approach is then validated using various benchmark models using real data sets from the education sector. Finally, our prediction method is discussed.

4.1. Analysis of Analysis

The case study based on the educational data set in this paper comes from students at Jinan Technical School. A sample questionnaire survey was conducted on computer applications, e-commerce, electrical automation, preschool education, and other professional students at Jinan Technical School, and a total of 442 questionnaires were issued and 417 valid questionnaires were retrieved. The questionnaire set up twelve questions of tasks, experimental scores, travel, interaction, number of problems, and admission. The score is 1 to 5 points, respectively. 70% of the samples are randomly selected from the sample as a training set, and the remaining 30% as a validation set.

Prior to data processing, the collected data is often pre-processed to improve the accuracy of predictions, and currently, mainly the types of abnormal data such as empty value data, repeat data, and abnormal data in the forecast are processed. The different types of abnormal data processing are as follows: the data is empty, and it is deleted directly when data is pre-processed. If the data is repeated, delete it. If there are abnormalities, remove the supplement.



(a) Importance order (b) Test set results (c) Error curve

Fig. 2 Result of random forest classification algorithm

We first analyze the characteristic data using random forest classification algorithms, setting the number of decision trees to 50 and the minimum leaf tree to 1, and sorting the characteristics by calculating the importance of the various characteristic data and output performance data, as shown in Fig. 2. The 1-12 in Fig. 2(a) are Paternal education, Maternal education, Number of hands raised, Homework time, Community activities, operation, Academic record, Economic situation, Psychological score, and Moral education achievement, Online class of time, Experimental score. According to Fig. 2(b), our random forest classification prediction accuracy reaches 96.29%. According to Fig. 2(c), which shows the trend of classification error rate with the increase of decision trees, we can see that when the number of decision trees increases to 40, the error rate decreases to a relatively low level.

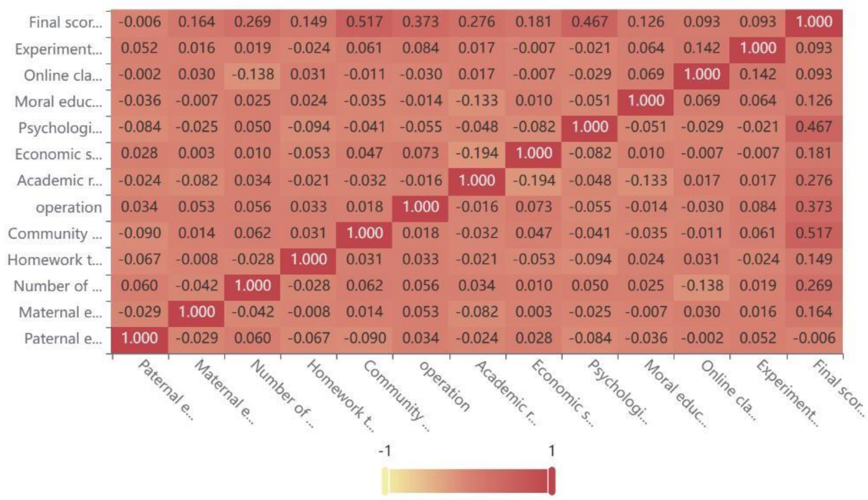


Fig. 3 Related Thermal Chart

According to Fig. 3, we can find that the seven characteristics of community activities, psychological scores, operation, academic record, Number of hands raised, economic situation, and maternal education have a high correlation with the final grade.

Based on the random forest importance ranking and the Pearson Correlation Coefficient, we screen out the seven characteristics of community activities, tasks, experimental scores, trips, psychological scores, problems, and admission scores as important characteristic data and reduce the complexity of follow-up algorithms through character selection methods. Based on the importance analysis of random forest classification and the Pearson correlation coefficient, we find that community activities play an extremely important role in the final practical achievement.

4.2. TLBO-BP Forecast Analysis

The TLBO-BP neural network methodology is based on the construction of a model for prediction of performance of middle-term students with practice, aimed at excavating the potential link between students participating in community activities and events, to the principle of early guidance and early effect. By conducting a prediction experiment on students' results at the Computer Applications, the experiment proves that there is a potential connection between participation in practical activities and students' practical results. At the same time, the student performance forecast model proposed has good prediction accuracy, practical usefulness, and can become an important part of improving the quality of teaching.

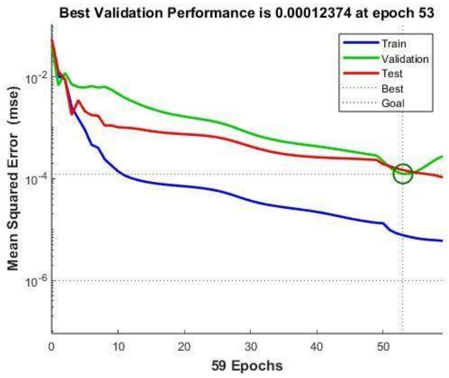
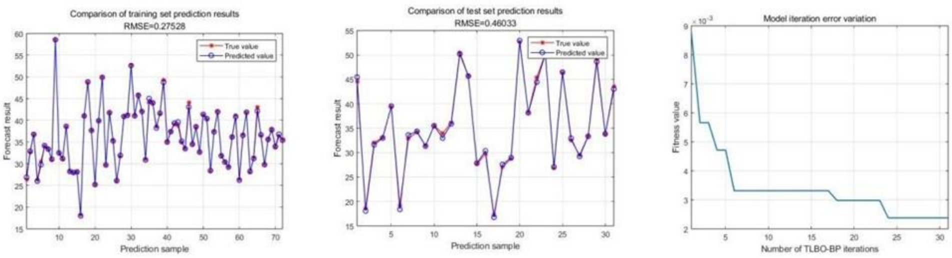


Fig. 4 BP neural network representation diagram

In the Fig. 4, the train curve is the performance of the MSE indicator of the BP training process in each generation, and the test curve is the performance of the MSE indicator of the BP test process in each generation. The Best dotted line shows that BP network training to the 53rd generation BP training results are the most ideal.



(a) Training set results (b) Prediction set results (c) Iterative error

Fig. 5 Graphical representation of the results of the TLBO-BP neural network

In Fig. 5, Fig. 5(a) refers to the training set prediction results for TLBO-BP with an RMSE error of 0.275. Fig. 5(b) refers to the test set prediction results for TLBO-BP with an RMSE error of 0.48. Fig. 5(c) shows the iterative error of the TLBO-BP neural network, which reaches its lowest error when iterating to generation 24.

Firstly, it can be seen from Fig. 4 and Fig. 5 that both training performance and testing performance are floating, which is different from the characteristics that the performance changes with the increase of training samples, indicating that BP algorithm does not need too many training samples and is less affected by the number of samples in the performance prediction. So that the performance prediction model based on BP algorithm can be constructed under the premise of less samples, and is competent for the performance prediction of fewer people. Secondly, the training error of the training samples does not exceed 0.5. Finally, for the test samples, the prediction error does not exceed 0.3. The excellent prediction effect can help teachers reform and innovate guidance strategies, promote students learning and continuous progress, and effectively improve the achievement of the goal of talent training.

We used GA-BP [13], ACO-BP and BP for prediction analysis on the same data set, and compared the results with TLBO-BP. The results comparison figure is as follows.

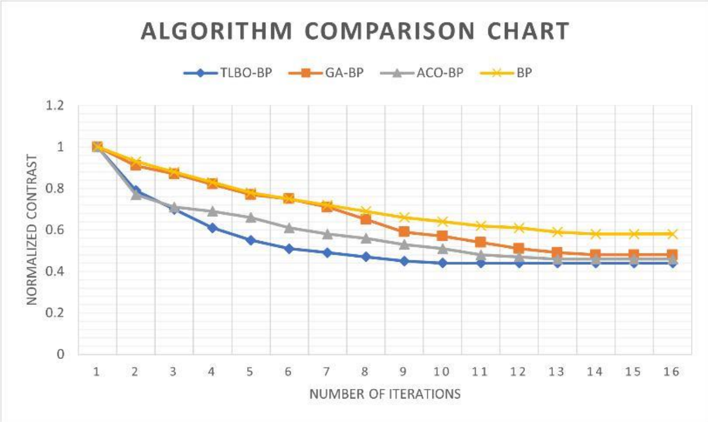


Fig.6 Comparison of prediction algorithms

In Fig. 6, the BP neural network is optimized by comparing different heuristic algorithms. We find that the BP neural network can be optimized by the TLBO algorithm, GA algorithm and ACO algorithm, but the TLBO algorithm can achieve faster optimization due to the small amount of hyperparameters and fast convergence.

It can be seen from the comparative experiments that the performance prediction model based on TLBO-BP algorithm can better use the scores of students participation in clubs to predict students 'practical scores, provide reference for teachers to carry out teaching guidance more effectively, and can play a greater role in talent training.

5. CONCLUSION

In this paper, we collect raw data by means of a questionnaire. Feature extraction was performed using random forest classification and Pearson correlation coefficient to exclude features that were not relevant to the final score. The initial weights and threshold in a BP neural network are optimized using the TLBO-BP algorithm, which avoids the instability of training results caused by stochastic initialization, and makes the full use of the advantages of a deep learning BP network and improves the prediction accuracy of the model. We compare the optimization of BP neural networks with different heuristics to prove that the optimization of TLBO algorithm is effective. In the case of insufficient data and many interfering factors, the prediction accuracy of 90% shows the good performance of the model.

In the future, we should analyze the impact of different educational models on students' student achievement, and it makes sense to collect data through more accurate data collection methods to take advantage of the role of advanced technology in the field of modern educational data mining.

AVAILABILITY OF DATA AND MATERIALS

The data cannot be shared; We are grateful to Jinan Institute of Technology for providing data and support for this study.

ACKNOWLEDGEMENTS

All errors remain mine. The author is thankful for all reviewers' comments, recommendations, and suggestions.

CONFLICT OF INTEREST

The authors received no financial support for the research, authorship, and/or publication of this article.

REFERENCE

[1] Elbadrawy, A. Polyzou, Z. Ren, M. Sweeney, G. Karypis, and H. Rangwala, "Predicting student performance using personalized analytics," Computer, vol. 49, pp. 61–69, 2016, doi: 10.1145/2959100.2959133.

[2] Z. Ren, X. Ning, A. Lan, and H. Rangwala, "Grade prediction based on cumulative knowledge and co-taken courses," in Proc. 12th Int. Conf.Educ. Data Mining, 2019, pp. 158–167.

[3] H. Nabizadeh, D. Gonçalves, S. Gama and J. Jorge, "Early Prediction of Students' Final Grades in a Gamified Course," in IEEE Transactions on Learning Technologies, vol. 15, no. 3, pp. 311-325, 1 June 2022, doi: 10.1109/TLT.2022.3170494.

[4] J. Dong, X. Liu, Z. Wang, L. Chen, F. Wang and J. Tang, "Research on Academic Early Warning Model Based on Improved SVM Algorithm," 2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), Dalian, China, 2022, pp. 1186-1191, doi: 10.1109/TOCS56154.2022.10016199.

[5] L. Song, "Sports Training Decision Support System and Realization Based on Data Mining and BP Neural Network," 2022 European Conference on Natural Language Processing and Information Retrieval (ECNLPIR), Hangzhou, China, 2022, pp. 1-4, doi: 10.1109/ECNLPIR57021.2022.00012.

[6] C. Hou, X. Yu, Y. Cao, C. Lai and Y. Cao, "Prediction of synchronous closing time of permanent magnetic actuator for vacuum circuit breaker based on PSO-BP," in IEEE Transactions on Dielectrics and Electrical Insulation, vol. 24, no. 6, pp. 3321-3326, Dec. 2017, doi: 10.1109/TDEI.2017.006475.

[7] Polyzou, A., Karypis, G. Grade prediction with models specific to students and courses. Int J Data Sci Anal 2, 159–171 (2016). https://doi.org/10.1007/s41060-016-0024-z

[8] Cortez, P. and A. M. Gonçalves Silva. "Using data mining to predict secondary school student performance." (2008).

[9] Polyzou, Agoritsa and George Karypis. "Feature extraction for classifying students based on their academic performance." *Educational Data Mining* (2018).

[10] Ren, Zhiyun et al. "Grade Prediction Based on Cumulative Knowledge and Co-taken Courses." *Educational Data Mining* (2019).

[11] M. R. Apriyadi, Ermatita, D. P. Rini and Yulnelly, "Implementation of Feature Selection Based on Particle Swarm Optimization and Genetic Algorithm on Support Vector Regression Algorithm to Predict Student Performance," 2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia, 2022, pp. 395-400, doi: 10.1109/ICIMCIS56303.2022.10017972.

[12] H. Nabizadeh, D. Gonçalves, S. Gama and J. Jorge, "Early Prediction of Students' Final Grades in a Gamified Course," in IEEE Transactions on Learning Technologies, vol. 15, no. 3, pp. 311-325, 1 June 2022, doi: 10.1109/TLT.2022.3170494.

[13] C. Mei, "A GA-BP Neural Network Model for Students Examination Score Prediction," 2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), Dalian, China, 2022, pp. 1212-1217, doi: 10.1109/TOCS56154.2022.10016064.

[14] H. Hu, J. Wu and X. Guan, "Research on ACO-BP Based Prediction Method of The Oilfield Production Stimulation Results," 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, China, 2020, pp. 240-243, doi: 10.1109/ICEIEC49280.2020.9152310.

[15] M. P. Neto and F. V. Paulovich, "Explainable Matrix - Visualization for Global and Local Interpretability of Random Forest Classification Ensembles," in IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 2, pp. 1427-1437, Feb. 2021, doi: 10.1109/TVCG.2020.3030354.

[16] W. Wang, W. Liu and H. Chen, "Information Granules-Based BP Neural Network for Long-Term Prediction of Time Series," in IEEE Transactions on Fuzzy Systems, vol. 29, no. 10, pp. 2975-2987, Oct. 2021, doi: 10.1109/TFUZZ.2020.3009764.
[17] M. Alizadeh, R. Rodriguez, J. Bauman and A. Emadi, "Optimal Design of Integrated Heat Pipe Air-Cooled System Using TLBO Algorithm for SiC MOSFET Converters," in IEEE Open Journal of Power Electronics, vol. 1, pp. 103-112,2020, doi: 10.1109/OJPEL.2020.2985700.
[18] S. I. Taheri, M. B. C. Salles and N. Kagan, "A new modified TLBO algorithm for placement of AVRs in distribution system," 2019 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America), Gramado, Brazil, 2019, pp. 1-6, doi: 10.1109/ISGT-LA.2019.8895471.
[19] L. Liu and S. Yu, "The Academic Warning Model Based on TLBO-BP Algorithm," 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Huangshan, China, 2018, pp. 65-70, doi: 10.1109/FSKD.2018.8687305.

AUTHORS

Huakun Hou is currently pursuing the masters degree with the School of Information Science and Engineering, Shandong Normal University. He current research interests include Education, artificial intelligence.

Xiaoke Deng is currently pursuing the masters degree with the School of Information Science and Engineering, Shandong Normal University. Her current research interests include Education, artificial intelligence.

Zekun Lu is currently pursuing the masters degree with the School of Information Science and Engineering, Shandong Normal University. He current research interests include UAV, IoT, edge computing.

Linbo Zhai (Member, IEEE) received his B. S. and M.S. degrees from School of Information Science and Engineering at Shandong University in 2004 and 2007, respectively. He received his Ph.D. degree from School of Electronic Engineering at Beijing University of Posts and Telecommunications in 2010. From then on, he worked as a teacher in Shandong Normal University. His current research interests include cognitive radio, crowd sourcing and distributed network optimization.

*corresponding author